Visualizing and Predicting Heart Diseases with an Interactive Dash Board

Team ID: PNT2022TMID19891

Faculty Mentor: Madumidha .S

Team Leader : Navaneethan C

Team Member : Renitha A

Team Member : Shobha N

Team Member : Ragul N

Team Member : Vibeshnan P

Project Objectives:

1. Age: displays the age of the individual.

2. Sex: displays the gender of the individual using the following format : 1 = male 0 = female

3. Chest-pain type: displays the type of chest-pain experienced by the individual using the following format : 1 = typical angina 2 = atypical angina 3 = non — anginal pain 4 = asymptotic

4. Resting Blood Pressure: displays the resting blood pressure value of an individual in mmHg (unit)

5. Serum Cholestrol: displays the serum cholesterol in mg/dl (unit)

6. Fasting Blood Sugar: compares the fasting blood sugar value of an individual with 120mg/dl. If fasting blood sugar > 120mg/dl then : 1 (true) else : 0 (false)

7. Resting ECG : displays resting electrocardiographic results 0 = normal 1 = having ST-T wave abnormality 2 = left ventricular hyperthrophy

8. Max heart rate achieved : displays the max heart rate achieved by an individual.

9. Exercise induced angina : 1 = yes 0 = no

10.ST depression induced by exercise relative to rest: displays the value which is an integer or float.

11.Peak exercise ST segment : 1 = upsloping 2 = flat 3 = downsloping
12.Number of major vessels (0–3) colored by flourosopy : displays the value as integer or float.

13.Thal : displays the thalassemia : 3 = normal 6 = fixed defect 7 = reversible defect

14.Diagnosis of heart disease : Displays whether the individual is suffering from heart disease or not : 0 = absence 1, 2, 3, 4 = present.

Data Analysis

Let us look at the people's age who are suffering from the disease or not.

Here, target = 1 implies that the person is suffering from heart disease and target = 0 implies the person is not suffering.

Variation of Age for each target class
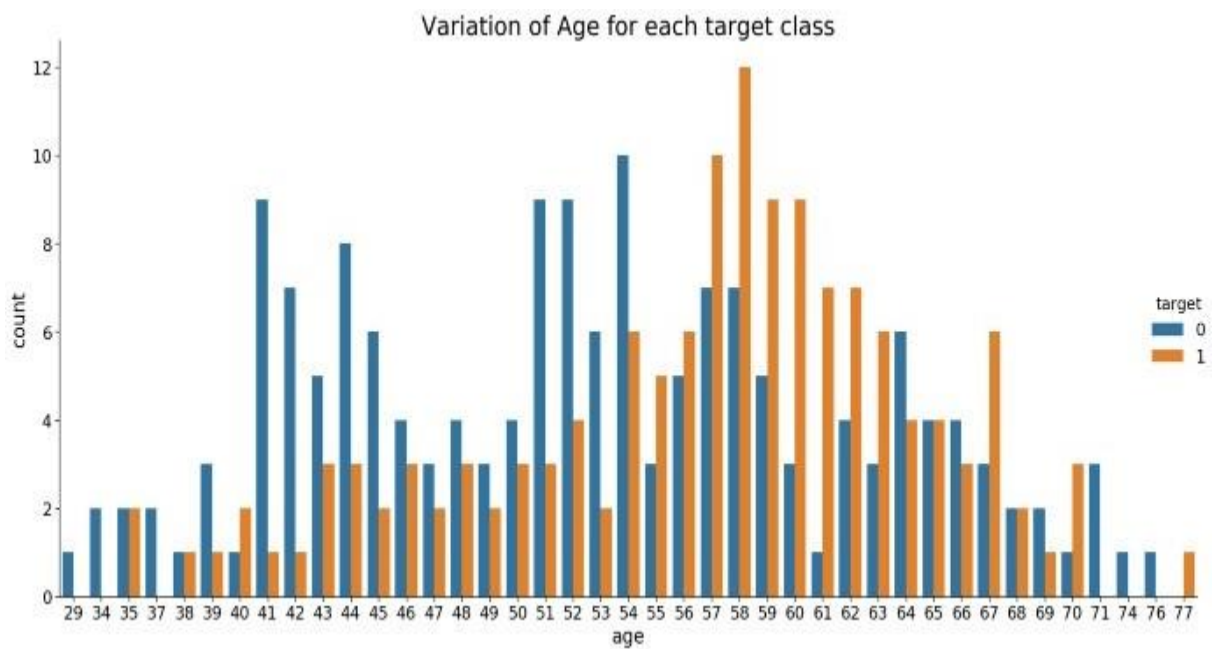
We see that most people who are suffering are of the age of 58, followed by 57. Majorly, people belonging to the age group 50+ are suffering from the disease.

Next, let us look at the distribution of age and gender for each target class.

Distribution of age vs sex with the target class
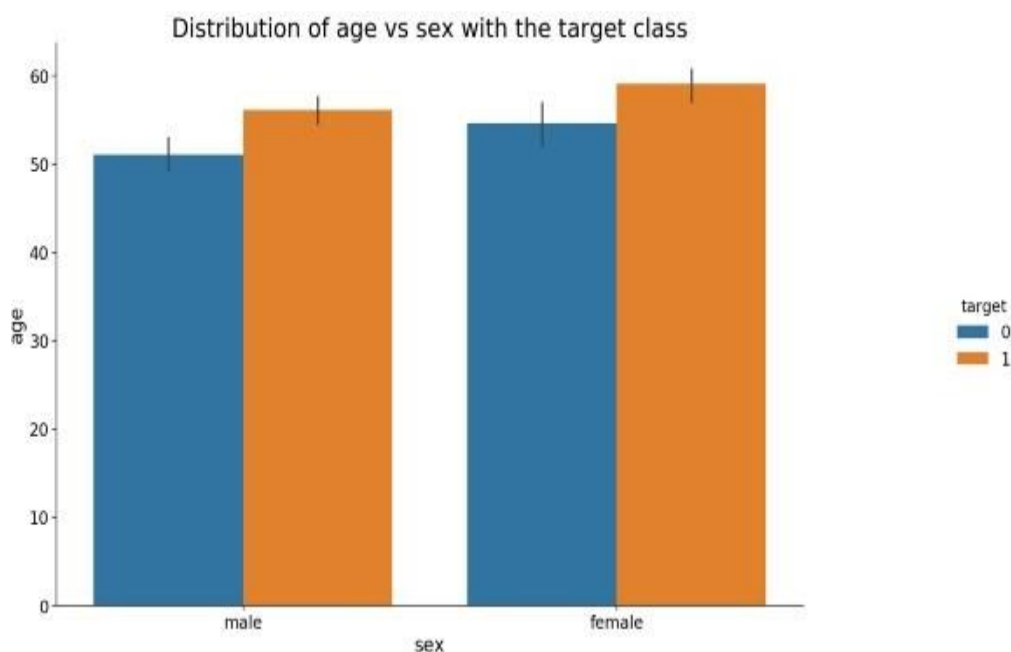
Data Pre-Processing

The dataset contains 14 columns and 303 rows.

Let us check the null values null values in each column of the data

```
Out[2]:
age            0
sex            0
cp             0
trestbps       0
chol           0
fbs            0
restecg        0
thalach        0
exang          0
oldpeak        0
slope          0
ca             4
thal           2
target         0
dtype: int64
```

We see that there are only 6 cells with null values with 4 belonging to attribute ca and 2 to thal.

As the null values are very less we can either drop them or impute them.

I have imputed the mean in place of the null values however one can also delete these rows entirely.

Now let us divide the data in the test and train set.

In this project, I have divided the data into an 80: 20 ratio. That is, the training size is 80% and testing size is 20% of the whole data.